

From: Randall K. Barry.

ALA-LC Romanization Tables; Transliteration Schemes
for non-Roman scripts approved by the Library of Congress
and the American Library Association.

Persian

Washington, D.C.: Library of Congress (1991)

Letters of the Alphabet

Initial	Medial	Final	Alone	Romanization
ا	ل	ل	ا	omit (see Note 1)
ب	ب	ب	ب	b
پ	پ	پ	پ	p
ت	ت	ت	ت	t
ث	ث	ث	ث	s
ج	ج	ج	ج	j
چ	چ	چ	چ	ch
ح	ح	ح	ح	h
خ	خ	خ	خ	kh
د	د	د	د	d
ذ	ذ	ذ	ذ	z
ر	ر	ر	ر	r
ز	ز	ز	ز	z
ش	ش	ش	ش	zh
س	س	س	س	s
ص	ص	ص	ص	sh
ض	ض	ض	ض	s
ظ	ظ	ظ	ظ	z
ع	ع	ع	ع	!
غ	غ	غ	غ	z
ف	ف	ف	ف	' (ayn)
ق	ق	ق	ق	gh
ک	ک	ک	ک	f
گ	گ	گ	گ	q
ل	ل	ل	ل	k (see Note 2)
م	م	م	م	g (see Note 3)
ن	ن	ن	ن	l
و	و	و	و	m
ه	ه	ه	ه	n
و	و	و	و	v (see Note 3)
ه	ه	ه	ه	h (see Note 4)
ی	ی	ی	ی	y (see Note 3)

Persian

Vowels and Diphthongs (see Note 5)

ā	a	ā, ā̇, ā̈	ā̇	ī
		(see Note 6)		
ū	u	ū, ū̇, ṻ	ū̇	aw
		(see Note 7)		
ī	i	ī, ī̇, ī̈	ī̇	ay

Notes

1. For the use of ʾ (alif) to support ʾ (hamzah) and ā̇ (maddah) see rule 1(a). For the romanization of ʾ and ā̇ see rules 4 and 5 respectively. For the use of ʾ to represent the long vowel romanized ā see the table of vowels and diphthongs, and rule 1(b).
2. Final ګ and ګ̇ (often written ګ and ګ̇) may have the form ګ̇, without the distinguishing upper stroke or strokes. The two letters are always distinguished in romanization.
3. For other values of و and ی see the table of vowels and diphthongs, and rules 2, 3, and 7.
4. ٲ (dotted ٲ) when used as an alternative to ٲ is romanized ٲ.
5. Vowel points are not printed on Library of Congress cards.
6. See rules 1(b) and 5.
7. See rule 3(d).

RULES OF APPLICATION

Letters Which May Be Romanized in Different Ways Depending on Their Context

1. ʾ (alif) is used:
 - (a) As a support for ʾ (hamzah) and ā̇ (maddah). In these cases it is not represented in romanization. See rules 4 and 5.
 - (b) To indicate the long vowel romanized ā. For the use of ʾ in tanvīn see rule 6.
2. و is used to represent:
 - (a) The consonant romanized v.

varzish
davā
sarv

ورزش
دوا
سرو

Silent و following خ is retained in romanization.

khvāstan
khvud

خواستن
خود

- (b) The long *u*-vowel (and short *u*-vowel in some monosyllables) is romanized *ū*.

dūr
chūn
tū

دور
چون
تو

- (c) The diphthong romanized *aw*.

Firdawsī

فردوسی

When the diphthong precedes a consonantal و, the combination is romanized *avv*. See rule 7.

و may be used as a support for ؤ (*hamzah*); in this case it is not represented in romanization. See rule 4.

3. ی is used to represent:

- (a) The consonant romanized *y*.

yār
siyāh
pāy

یار
سیاه
پای

- (b) The long vowel romanized *ī*.

Īrān
qālī

ایران
قالی

- (c) The diphthong romanized *ay*.

ayvān
ray

ایوان
ری

- (d) The final long vowel romanized *ā*.

Muṣṭafā

مصطفی

For the use of ی as a mark of *izāfah* see rule 8(c).

ی in the medial forms ـی, ـی, without dots, may be used as a support for ؤ (*hamzah*); in this case ی is not represented in romanization. See rule 4 below.

Orthographic Symbols Other Than Letters and Vowel Signs

The signs listed below are frequently omitted in Persian writing and printing; their presence must then be inferred. They are represented in romanization according to the following rules:

4. ؤ (hamzah)

(a) When initial, ؤ is not represented in romanization.

(b) When medial or final, ؤ is romanized ' (alif) except as noted in (c) and (d) below.

mu'assir

مؤثر

khulafā'

خلفاء

pā'īn

پائین

(c) When used as a mark of *izāfah*, ؤ is romanized -'i.

astānah-'i dar

استانه در

(d) When used to mark the indefinite article, ؤ is romanized 'i.

khānah'i

خانه

5. ٲ (maddah)

(a) Initial ٲ is romanized ā

āb

Kullīyat al-Ādāb

آب
کلیه الآداب

(b) Medial ٲ, when it represents the phonetic combination 'ā, is so romanized.

ma'āsir

مآثر

Daryā'ābādī

دریاآبادی

(c) ٲ is otherwise not represented in romanization.

gardāvarandah

گردآورنده

6. *Tanvīn* (written ٲ, ٲ, ٲ, ٲ), which occurs chiefly in Arabic words, is romanized *un*, *in*, *an*, and *an*, respectively.

7. ٲ (shaddah or *tashdīd*) is represented by doubling the letter or digraph concerned.

khurram

خرم

avval

اول

bachchah

بچه

Khayyām

خیام

Note the exceptional case where ڤ is written over و and ی to represent the combination of long vowel plus consonant.

nashriyāt	نشریات
qūvah	قوه

Grammatical Structure as It Affects Romanization

8. *Izāfah*. When two words are associated in the relation known as *izāfah*, the first (the *muzāf*) is followed by an additional letter or syllable in romanization. This is added according to the following rules:

- (a) When the *muzāf* bears no special mark of *izāfah*, it is followed by -i.

dar-i bāgh	در باغ
qālī-i Īrān	قالی ایران
khānah-i buzurg	خانه بزرگ

- (b) When the *muzāf* is marked by the addition of ة, it is followed by -i.

qālī-ī Īrān	قالی ایران
khānah-ī buzurg	خانه بزرگ

- (c) When the *muzāf* is marked by the addition of ی, it is followed by -yi.

rū-yi zamīn	روی زمین
Daryā-yi Khazar	دریای خزر
khānah-yi buzurg	خانه بزرگ

- (d) *Izāfah* is represented in romanization of personal names only when expressly indicated in the Persian script.

Affixes and Compounds

9. Affixes.

- (a) When the affix and the word with which it is connected grammatically are written separately in Persian, the two are separated in romanization by a single prime ('). See also 12(b) below.

	khānah' hā	خانه ها
	khānah' am	خانه ام
	khānah' ī	خانه ای
	mī' ravam	می روم
but	mī ravam	میروم
	bih' gū	به گو
	bar' rasihā	بررسیها
	Kāzim' zādah	کاظم زاده
but	Kāzimzādah	کاظمزاده

Persian

- (b) The Arabic article *al* is separated by a hyphen, in romanization, from the word to which it is prefixed.

dār al-mu'allimīn
'Abd al-Ḥusayn

دار المعلمین
عبد الحسین

10. Compounds. When the elements of a compound (except a compound personal name) are written separately in Persian, they are separated in romanization by a single prime ('). See also 12(b) below.

marīẓ/khānah
but marīẓkhānah
Shāh/nāmah
but Shāhnāmah

مریض خانه
مریضخانه
شاه نامه
شاهنامه

Note the treatment of compound personal names:

Ghulām 'Alī
Shāh Jahān
Ibn Abī Ṭālib

غلام علی or غلامعلی
شاه جهان or شاهجهان
ابن ابی طالب or ابن ابیطالب

Orthography of Persian in Romanization

11. Capitalization.

- (a) Rules for the capitalization of English are followed, except that the Arabic article *al* is lower cased in all positions.
- (b) Diacritics are used with both capital and lower case letters.

12. The single prime (') is used:

- (a) To separate two letters representing two distinct consonantal sounds, when the combination might otherwise be read as a digraph.

marz'hā

مرزها

- (b) To mark the use of a letter in its final form when it occurs in the middle of a word. See also rules 9(a) and 10 above.

rāh'hā

Qāyim'maqāmī

Bih'āzīn

راهها

قایم مقامی

به آذین

13. Foreign words in a Persian context, including Arabic words, are romanized according to the rules for Persian. For short vowels not indicated in the script, the Persian vowels nearest the original pronunciation of the word are supplied in romanization.
14. Dictionaries.

In romanizing Persian, the Library of Congress has found it necessary to consult dictionaries as an appendage to the romanization tables, primarily for the purpose of supplying vowels. For Persian, the principle dictionary consulted is:

M. Mu'in. *Farhang-i Fārsī-i mutavassit*.

SPECIAL CHARACTERS AND CHARACTER MODIFIERS IN ROMANIZATION

<i>Special Characters</i>	<i>Name</i>	<i>USMARC hexadecimal code</i>
/	soft sign (prime)	A7
,	alif	AE
'	ayn	B0
<i>Character Modifiers</i>	<i>Name</i>	<i>USMARC hexadecimal code</i>
ó	acute	E2
ō	macron	E5
◌̣	dot below	F2
◌̣̣	double dot below	F3
◌̣̣̣	underscore	F6

Pushto

Letters of the Alphabet

Initial	Medial	Final	Alone	Romanization
ا	ل	ل	ا	omit (see Note 1)
ب	پ	ب	ب	b
پ	ت	پ	پ	p
ت	ث	ت	ت	t
ث	ج	ث	ث	t̤
ج	چ	ج	چ	ʒ
چ	ح	چ	ح	j
ح	خ	ح	خ	ch
خ	ځ	خ	ځ	h̥
ځ	ښ	ځ	ښ	s̥
ښ	د	ښ	د	z̥
د	ډ	د	ډ	kh
ډ	ر	ډ	ر	d
ر	ړ	ړ	ړ	ɖ
ړ	ز	ړ	ز	z
ز	ږ	ږ	ږ	r
ږ	س	ږ	س	r̥
س	ش	ش	ش	z
ش	ښ	ښ	ښ	zh
ښ	ښ	ښ	ښ	zh̥
ښ	ښ	ښ	ښ	ʒ
ښ	ښ	ښ	ښ	s
ښ	ښ	ښ	ښ	sh
ښ	ښ	ښ	ښ	sh̥
ښ	ښ	ښ	ښ	ʂ
ښ	ښ	ښ	ښ	z̥
ښ	ښ	ښ	ښ	t̤
ښ	ښ	ښ	ښ	t̤
ښ	ښ	ښ	ښ	z̥
ښ	ښ	ښ	ښ	' (ayn)
ښ	ښ	ښ	ښ	gh
ښ	ښ	ښ	ښ	f
ښ	ښ	ښ	ښ	q
ښ	ښ	ښ	ښ	k
ښ	ښ	ښ	ښ	g
ښ	ښ	ښ	ښ	g
ښ	ښ	ښ	ښ	l
ښ	ښ	ښ	ښ	m
ښ	ښ	ښ	ښ	n
ښ	ښ	ښ	ښ	ŋ

نم	نم	نم	نم	ټ
و	و	و	و	w (see Note 2)
ه	ه	ه ، ء	ه ، ء	h
ي	ي	ي ، ي	ي ، ي	y (see Note 3)

Vowels and Diphthongs (see Note 4)

Romanization	Initial	Medial	Final
a (see Note 5)	آ	ا	ا
u	أ	ؤ	ؤ
i	إ	ي	ي
ā	آ	ا	ا
á	-	-	(see Note 6) ي
ū	أ	ؤ	ؤ
ī	إ	ي	ي
o	او	و	و
e	اي ، اپ	ي ، ي	ي ، ي ، ي ، ي
aw	آو	ؤ	ؤ
ay	آي	ي	ي ، ي
ə	-	-	ي ، ي ، ع

Notes

1. For the use of ا (alif) see the table of vowels and diphthongs, and rules 1-4.
2. For other uses of و see the table of vowels and diphthongs, and rule 5.
3. For other forms and used of ي see the table of vowels and diphthongs, and rules 6 and 7.
4. Vowel points are rarely found in Pushto writing and printing. Vowels not indicated in the script are supplied in romanization.
5. In addition to the three short vowels recognized by the orthography of all languages using Arabic script, Pushto possesses a short, central vowel for which the sign ا (differentiated from ا) has sometimes been used. Both signs, whether written or inferred, may be romanized a. When this central vowel is indicated by ʾ (hamzah), the vowel is romanized ə; the hamzah in this case is not represented in romanization. See rule 8(c) and (d). When it is desired for any reason to show the presence of a central vowel in a particular word, the romanization ə may be used.
6. See rule 7(a).
7. See rule 7(b).

RULES OF APPLICATION

Letters Which May Be Romanized in Different Ways Depending on Their Context

1. ^l (*alif*) is used:

(a) To indicate the presence of any of the short vowels *a, u, i*.
 atah اته
 ulas الس
 in'ām انعام

(b) To represent the long vowel romanized *ā*. When so used, ^l is usually written with [̄] (*maddah*).
 ādam آدم

(c) To represent, in combination with و or ی :

(1) The long vowels romanized *ū* and *ī* respectively.
 ūṣh اوبښ
 Īrān ايران

(2) The long vowels romanized *o* and *e* respectively.
 os اوس
 eṣhal ايښل

(3) The diphthongs romanized *aw* and *ay* respectively.
 awwal اول
 ayyām ايام

2. ^l (*alif*), when medial or final, is used to represent the long vowel romanized *ā*.
 bābā بابا

3. ^l (*alif*), when final, sometimes represents the combination romanized *an*. This value is confined to a few words of Arabic origin. See rule 10.

4. ^l (*alif*), when used to support ء (*hamzah*), is not represented in romanization; see rule 8.

5. و is used to represent:

(a) The consonant romanized *w*.
 wror ورور

In Persian words, silent و following ځ is retained in romanization.
 khwushhāl خوشحال

(b) The long vowel romanized *ū*. See also rule 1(c1).
 nūm نوم

(c) The long vowel romanized *o*. See also rule 1(c2).
 kor کور

- (d) The diphthong romanized *aw*. See also rule 1(c3).
yaw يو

- (e) The short vowel romanized *u*. Such cases are rare in Pushto orthography.
u او

Note: و when used to support ء (*hamzah*) is not represented in romanization. See rule 8.

6. ي is used to represent:

- (a) The consonant romanized *y*.
yaw يو

- (b) The long vowel romanized *ī*. See also rule 1(c1)
mīnah مینه

- (a) The long vowel romanized *e*. See also rule 1(c2).
der ډیر ، ډیر
dre درې ، درې ، درے

- (c) The diphthong romanized *ay*. See also rule 1(c3).
shay شای

Note: Medial ی when used to support ء (*hamzah*) is usually written without dots (ٹ ، ٹ); it is not represented in romanization. See also rule 8(b) and (c)

7. ی when final has two special uses. It may represent:

- (a) The long vowel romanized *ā*. This use is confined to word of Arabic origin.
Muṣṭafā مصطفیٰ

- (b) The diphthong romanized *ay*. When used with this value, ی usually has the form غ in Peshawar, ی in Afghanistan. See also rule 8(c).

doḡay دوغی ، دوغ

Orthographic Symbols Other Than Letters and Vowel Signs

The signs listed below are often omitted in Pushto writing and printing; their presence must be inferred. They are represented in romanization according to the following rules:

8. ء (*hamzah*)

- (a) In initial position, ء is not represented in romanization.

- (b) In words of Arabic origin, ء when medial or final is romanized ' (*alif*).

su'āl سؤال
dā'imī دائمی
mas'alah مساله

Pushto

- (c) When written over final ی or پ, ء indicates the diphthong romanized *ay*. See rule 7(b).
ghwārāy غوارای
- (d) In the orthography of Peshawar, ء when written over final ټ indicates the short vowel romanized *a*. See the table of vowels and diphthongs, footnote 7.
naḥ نه
9. ̄ (maddah) is written over ټ (alif) to indicate the long vowel romanized *ā*.
ākhir آخر
10. Tanwīn (written ّ) occurs in words of Arabic origin and represents a final syllable which is romanized *an*. Tanwīn is usually associate with ټ (alif), which is not represented in romanization. See rule 3.
fīṭṭān فطرتا
11. ّ (tashdīd) is represented by doubling the letter or digraph concerned.
awwāl اول

Grammatical Structure as It Affects Romanization

12. *Izāfah*. When two words are associated in the Persian construction known as *izāfah*, the first (the *muẓāf*) is followed by an additional letter or syllable in romanization. This is added according to the following rules:
- (a) When the *muẓāf* bears no special mark of *izāfah*, it is followed by *-i*.
dars-i 'ibrat درس عبرت
ṣāḥib-i mubārak صاحب مبارک
- (b) When the *muẓāf* is marked by the addition of ء (*hamzah*), it is followed by *-i*.
- (c) When the *muẓāf* is marked by the addition of ی, it is followed by *-yi*.

Affixes and Compounds

13. Affixes.

When the affix and the word with which it is connected grammatically are written separately in Pushto, the two are separated in romanization by a single prime (').

nā'ashnā نلشنا
war'astawī وراستوی

Note 1: The particle ډ (*da*), meaning "of", is always written as a separate word in romanization.
da ḍaḥ د ده

Note 2: The Arabic article *al* is separated from what follows by a hyphen. The *l* of the article is retained in romanization, regardless of whether or not this *l* is assimilated in pronunciation to the sound of the following letter.

'Abd al-Rashīd عبد الرشید
Faḍl al-Ḥaqq فضل الحق
Niẓām al-Dīn نظام الدین

14. Compounds.

The elements of a compound (except a compound personal name) are separated in romanization by a single prime (').

	kitāb'khānah	کتابخانه
but	marīẓkhānah	کتابخانه

Orthography of Pushto in Romanization

15. Capitalization.

(a) Rules for the capitalization of English are followed, except that the Arabic article *al* is lower cased in all positions.

(b) Diacritics are used with both capital and lower case letters.

16. Foreign words in a Pushto context, including Arabic, Persian, and Urdu words, are romanized according to the rules for Pushto. For vowels not indicated in the script, the Pushto vowels nearest those in the original pronunciation of the word are supplied in romanization.

Yūnīwarsiṭī Buk Ejānsī	یونیورسٹی بک ایجنسی
(not University Book Agency)	
Dārmastatar	دارمستتر
(not Darmesteter)	

SPECIAL CHARACTERS AND CHARACTER MODIFIERS IN ROMANIZATION

<i>Special Characters</i>	<i>Name</i>	<i>USMARC hexadecimal code</i>
'	soft sign (prime)	A7
ا	alif	AE
آ	ayn	B0
<i>Character Modifiers</i>	<i>Name</i>	<i>USMARC hexadecimal code</i>
◌̊	acute	E2
◌̄	macron	E5
◌̇	dot above	E7
◌̈	dot below	F2
◌̈̈	double dot below	F3
◌̲	underscore	F6

Urdu

(in Arabic Script)

Letters of the Alphabet

Initial	Medial	Final	Alone	Romanization
ا	ل	ل	ا	omit (see Note 1)
ب	ب	ب	ب	b
پ	پ	پ	پ	p
ت	ت	ت	ت	t
ٹ	ٹ	ٹ	ٹ	! !
ث	ث	ث	ث	! !
ج	ج	ج	ج	j
چ	چ	چ	چ	c
ح	ح	ح	ح	h
خ	خ	خ	خ	kh
د	د	د	د	d
ڈ	ڈ	ڈ	ڈ	ḍ
ذ	ذ	ذ	ذ	ḍ
ر	ر	ر	ر	z
ز	ز	ز	ز	r
س	س	س	س	! !
ش	ش	ش	ش	! !
ص	ص	ص	ص	z
ض	ض	ض	ض	zh
ط	ط	ط	ط	s
ظ	ظ	ظ	ظ	sh
ع	ع	ع	ع	ṣ
غ	غ	غ	غ	z
ف	ف	ف	ف	! !
ق	ق	ق	ق	z
ک	ک	ک	ک	' (ayn)
گ	گ	گ	گ	gh
ل	ل	ل	ل	f
م	م	م	م	q
				k
				g
				l
				m

ن	ن	ن	ن	n
و	و	و	و	n (see Note 2)
و	و	و	و	v
و	و	و	و	h
و	و	و	و	t (see Rule 10)
و	و	و	و	y (see Note 3)

Digraphs Representing Urdu Aspirates (see Note 4)

	Value
bh	بھ
ph	پھ
th	تھ
ṭh	ٹھ
jh	جھ
ch	چھ
dh	دھ
ḍh	ڈھ
ṛh	رھ
kh	کھ
gh	گھ

Urdu Vowels and Diphthongs (see Note 5)

	Value
a	ا
u	و
i	ی
ā	آ
á	آی , آئی
ū	وو
ī	ی
o	و
e	ے , ای
au	او
ai	ای , آئی

Notes to the Tables

1. For the use of ʾ (*alif*) to support ʰ (*hamzah*) and ̣ (*maddah*) see rules 1 and 2, respectively. For the romanization of ʰ by ' (*alif*) see rule 12. For other orthographic uses of ʾ see rules 3-4.
2. For the distinction between ن and و see rule 6.
3. For the distinction between ی and ے see rule 11(c) and (e).
4. For the form of the letter ھ in these digraphs, see rule 9.
5. Vowel points are used sparingly, and for romanization must be supplied from a dictionary.

RULES OF APPLICATION

Letters Which May Be Romanized in Different Ways Depending on Their Context

1. ^l (*alif*), و and ی are used to support ء (*hamzah*); see rule 12. When so used, these letters are not represented in romanization.
2. ^l (*alif*) is used to support َ (maddah); see rule 13. When so used, it is not represented in romanization.
3. ^l (*alif*) is used after a consonant to indicate the long vowel romanized *ā*.

rāj
karnā

راج
کرنا

In some words of Arabic origin this *alif* appears as a superscript letter over ی representing the *alif maqṣūrah*.

da‘vā

دعویٰ

The *alif* is sometimes omitted in writing. It is always represented in romanization.

‘Abdurrahmān عبد الرحمن، عبد الرحمان

When the long vowel *ā* is initial, it is written ^l. See rule 13(a).

4. ^l (*alif*) may be used as an orthographic sign without phonetic significance. In these cases it is not represented in romanization. See rule 16.

‘ilman

علما

5. ط appears as a superscript letter over ت، د، and ر when the latter represent the cerebral sounds romanized *t*, *d*, and *r*, respectively.

6. Regardless of pronunciation, undotted forms of the letter ن are romanized *n* and dotted forms are romanized *ṇ*.

jahān

جہاں

7. و is used:

- (a) To represent the consonant sound romanized *v*.

dev
vujūd

دیو
وجود

In some words of Persian origin this consonant, though written, has ceased to be pronounced. It is retained in romanization.

khvīsh

خویش

- (b) To represent the long vowel romanized *ū*.

ūkh
Urdū

اوخ
اردو

- (c) To represent the long vowel romanized *o*.

os
dost

اوس
دوست

For the romanization of the conjunction و as *o* see rule 19.

- (d) To represent the diphthong romanized *au*.

aur
qaumī

اور
قومی

- (e) To support ء (*hamzah*). See rule 12.

For the use of ّ (*shaddah*) with و see rule 14.

8. ◌ is used to represent the consonantal sound romanized *h*.

ham
gāh

هم
گاہ

Final ◌, though often not pronounced, is normally retained in romanization.

kih
guldastah

کہ
گلدستہ

Exception is made in the case of words whose final syllable ends in an aspirated consonant. When final ◌ is added to the letter ہ in this position, it is not represented in romanization.

mukh

مکھہ

9. ◌ (usually written in the form ہ) is used to represent the aspirated element of the sounds romanized *bh, ph, th, ṭh, jh, ch, dh, ḍh, ṛh, kh, gh*.

phūl
acchā

پھول
اچھا

For the writing and romanization of words ending in an aspirated consonant, see rule 8.

10. ّ and ت, which are sometimes used interchangeably, are both romanized *t*.

ḥikmat

حکمت , حکمة

11. ی (often written ے) is used:

- (a) To represent the consonant romanized *y*.

siyāsat
dayā

سیاست
دیا

Urdu

- (b) To represent the long vowel romanized *ī*.

taṣvīr
īshvār

تصویر
ایشوار

- (c) To represent the long vowel romanized *e*.

sher
nevā

شیر
نیوا

When ی with this value is final, the form ے generally replaces ی.

se
laṛke

سے
لڑکے

- (d) To represent the long vowel romanized *ā*. See rule 3.

da'vā
'uqbā

دعوی
عقبی

- (e) To represent the diphthong romanized *ai*.

maidān
bail

میدان
بیل

When ی with this value is final, it is sometimes written ے.

hai

ہے

- (f) To support ء (*hamzah*). In this position ی is usually undotted. See rule 12.

For the use of ّ (*shaddah*) with ی see rule 14.

For the use of ی in a *muṣṣaf* see rule 17.

Romanization of Orthographic Symbols Other Than Letters and Vowel Signs

Although vowel signs are frequently omitted in printed texts, they are always taken into consideration in romanization. The rules for other symbols vary.

12. ء (*hamzah*)

- (a) In initial position ء is not represented in romanization.

- (b) In medial and final position, when ء represents a consonant, it is romanized ' (*alif*).

mu'min
li'e
bhā'ī

مؤمن
لئے
بھائی

- (c) When ة represents the connective syllable joining a *muzāf* to what follows, it is romanized -*yi*. See rule 17.

malikah-yi Inglisṭān ملکہ انگلستان

13. ̣ (maddah)

- (a) At the beginning of a word, or following the Arabic article ال, ̣ is romanized ā.
āb آب

- (b) At the beginning of a syllable within a word, ̣ is romanized 'ā.
mir'āt مرآت
Qur'ān قرآن

- (c) ̣ is otherwise omitted in romanization.

14. ̣̣ (shaddah or tashdīd) indicates the doubling in pronunciation of the letter over which it is written. It is represented in romanization by doubling the letter or digraph concerned.

caccā چچا
khaṭṭā کھٹا
makkhī مکھی

When ̣̣ occurs over و and ی, these letters are regarded as representing consonants. They are romanized vv and yy, respectively.

quvvat قوت
sayyid سید
Zakariyyā زکریا

15. ̣̣̣ (sukūn or jazm) indicates the absence of a vowel following the letter over which it is written. It is not represented in romanization.

16. Tanvīn (written ̣̣̣, ̣̣̣̣, ̣̣̣̣̣ (̣̣̣̣̣̣)) is romanized *un*, *in*, *an*, respectively, when it occurs in a word or expression borrowed from Arabic. Otherwise it is not represented in romanization.

fauran فوراً

Romanization as Affected by Grammatical Structure

17. Izāfat.

- (a) When the *muzāf*, the first of two words in the grammatical relationship known as *izāfah*, ends in a consonant, -*i* is added to it in the romanization.

tārīkh-i Hindūstān تاریخ ہندوستان

- (b) When the *muzāf* ends in a vowel or in silent *o*, *-yi* is added.

daryā-yi shor	دریای شور
zabān-i Urdū-yi mu'allā	زبان اردو معلی
malikah-yi Inglisān	ملکہ انگلستان

For the use of *o* (*hamzah*) to indicate the *muzāf*, see rule 12(c).

18. The Arabic article *al* is romanized differently depending on the letters and context with which it is associated.

- (a) When it is prefixed to a word beginning with a "moon letter" (*ا، ی، ب، ج، ح، خ، ع، غ، ف، ق، ک*) it is romanized *al*.

al-Qur'ān القرآن

- (b) When it is prefixed to a word beginning with a "sun letter" (*ظ، ل، ن، ت، ث، د، ذ، ر، س، ش، ص، ض، ط*) the *l* of the article is replaced in romanization by the same letter, or digraph, as that which begins the following word.

as-sijill السجل

- (c) When it occurs before the second element in a name, the vowel of the article is replaced by the final vowel of the preceding word.

'Abdul'azīz	عبد العزیز
'Abdurrashīd	عبد الرشید
Abūlfazl	ابو الفضل
Zūlqarnain	ذو القرنین
Faḡlullāh	فضل اللہ

19. The conjunction *و*, when used to join two closely associated members of a phrase, is romanized *o*.

māl o asbāb مال و اسباب

Otherwise *و* is romanized *va*.

20. Rules for the capitalization of English are followed, except that the Arabic article *al* is lowercased in all positions.

21. The macron is used with both capital and lowercase letters.

22. The hyphen is used:

- (a) To connect a *muzāf* with the following vowel or syllable. See rule 17.

- (b) To connect the Arabic article *al* with the following word. See rule 18.

23. Foreign words in an Urdu context, including Arabic and Persian words, are romanized according to the rules for Urdu.

جناب ہیڈ ماسٹر صاحب گورنمنٹ ہائی اسکول
Janāb-i Heḍ Māṣṭar ṣāhib-i Gavarnmanṭ Hā'ī Iskūl

For short vowels not indicated in the script, the Urdu vowels nearest the original pronunciation of the word concerned are supplied in romanization.

24. A quotation in another language using the Arabic script is romanized according to the rules for the language concerned.

SPECIAL CHARACTERS AND CHARACTER MODIFIERS IN ROMANIZATION

<i>Special Characters</i>	<i>Name</i>	<i>USMARC hexadecimal code</i>
'	alif	AE
‘	ayn	B0
<i>Character Modifiers</i>	<i>Name</i>	<i>USMARC hexadecimal code</i>
á	acute	E2
ā	macron	E5
◌̣	dot below	F2
◌̤	double dot below	F3
◌̥	underscore	F6

Index to Languages

This index lists the names of languages covered by romanization schemes in this publication. After each name is an indication of where the table for that language can be found. Some tables are applied to more than one language (e.g., the table for Hindi is used also for Awadhi, Bihari, Braj, Kashmiri in Devanagari script, Maithili, Pahari, and Rajasthani). In these cases, an instruction to use a specific romanization table is given. A table is generally named for the language for which it was first developed. See references from alternate names and spellings of languages have been included in the index when they differ significantly from the preferred form.

<i>Language</i>	<i>page</i>
Abaza	see Non-Slavic Languages 115
Abkhaz	see Non-Slavic Languages 115
Adygei	see Non-Slavic Languages 115
Aisor	see Non-Slavic Languages 115
Altai	see Non-Slavic Languages 115
Amharic	2
Arabic	4
Armenian	14
Assamese	16
Avar	see Non-Slavic Languages 116
Awadhi	use Hindi 54
Azerbaijani	see Non-Slavic Languages 116
Balkar	see Non-Slavic Languages 116
Bashkir	see Non-Slavic Languages 116
Belorussian	18
Bengali	20
Bhutan	see Tibetan 192
Bihari	use Hindi 54
Braj	use Hindi 54
Bulgarian	22
Burmese	24
Buryat	see Non-Slavic Languages 117
Byelorussian	see Belorussian 18
Cambodian	see Khmer 76
Chechen	see Non-Slavic Languages 117
Chinese	27
Chukchi	see Non-Slavic Languages 118
Church Slavic	40
Chuvash	see Non-Slavic Languages 118
Coptic	see Hebrew and Yiddish 53

Index to Languages

<i>Language</i>	<i>page</i>
Croatian	see Serbian and Macedonian 166
Cyrillic script	see Belorussian, Bulgarian, Church Slavic, Non-Slavic Languages, Russian, Serbian and Macedonian, or Ukrainian
Dargwa	see Non-Slavic Languages 118
Devanagari script	see Bengali, Hindi, Marathi, or Sanskrit and Prakrit
Divehi 43
Dogri	use Punjabi 152
Dungan	see Non-Slavic Languages 119
Eskimo	see Non-Slavic Languages 119
Even	see Non-Slavic Languages 119
Evenki	see Non-Slavic Languages 119
Farsi	see Persian 145
Gagauz	see Non-Slavic Languages 119
Galig	see Mongolian 108
Georgian 46
Gilyak	see Non-Slavic Languages--Nivkh 128
Gipsy	see Non-Slavic Languages--Romany 128
Greek 48
Gujarati 50
Gurmukhi script	see Punjabi 152
Hebrew 52
Hindi 54
Ingush	see Non-Slavic Languages 119
Inuit	see Non-Slavic Languages--Eskimo 119
Japanese 57
Kabardian	see Non-Slavic Languages 120
Kalmyk	see Non-Slavic Languages 120
Kannada 70
Karachay	see Non-Slavic Languages 120
Karachay-Balkar	see Non-Slavic Languages 120
Karakalpak	see Non-Slavic Languages 120
Karelian	see Non-Slavic Languages 121
Kashmiri 72
Kazakh	see Non-Slavic Languages 121
Khakass	see Non-Slavic Languages 121
Khanty	see Non-Slavic Languages 122
Khmer 76
Kirghiz	see Non-Slavic Languages 123
Komi	see Non-Slavic Languages 123
Komi-Permyak	see Non-Slavic Languages 123
Konkani	use Kannada 70 or Marathi 106

<i>Language</i>	<i>page</i>
Korean	79
Koryak	see Non-Slavic Languages 124
Kumyk	see Non-Slavic Languages 124
Kurdish (in Perso-Arabic script)	94
Lahada	use Punjabi 152
Lak	see Non-Slavic Languages 124
Lao	96
Lapp	see Non-Slavic Languages 125
Lepcha	100
Lezghian	see Non-Slavic Languages 125
Limbu	102
Lithuanian	see Non-Slavic Languages 126
Macedonian	see Serbian and Macedonian 166
Maithili	use Hindi 54
Malayalam	104
Maldivian	see Divehi 43
Manipuri	use Bengali 20
Mansi	see Non-Slavic Languages 126
Marathi	106
Mari	see Non-Slavic Languages 126
Moldavian	see Non-Slavic Languages 127
Molodtsov	see Non-Slavic Languages--Komi 123
Mongolian (in Cyrillic script)	see Non-Slavic Languages 127
Mongolian (in Vertical script)	108
Moplah	110
Mordvin	see Non-Slavic Languages 127
Nanai	see Non-Slavic Languages 127
Nenets	see Non-Slavic Languages 128
Nepali	use Hindi 54
Newari	use Hindi 54
Nivkh	see Non-Slavic Languages 128
Nogai	see Non-Slavic Languages 128
Non-Slavic Languages (in Cyrillic script)	114
Oriya	132
Ossetic	see Non-Slavic Languages 128
Ottoman Turkish	134
Pahari	use Hindi 54
Pali (in various scripts)	140
Panjabi	see Punjabi 152
Permyak	see Non-Slavic Languages--Komi-Permyak 123
Persian	145
Prakrit	see Sanskrit and Pankrit 162

Index to Languages

<i>Language</i>	<i>page</i>
Punjabi	152
Pusho	154
Rajasthani use Hindi	54
Romanian see Non-Slavic Languages--Moldavian	127
Romany see Non-Slavic Languages	128
Russian	160
Sanskrit	162
Santali	164
Selkup see Non-Slavic Languages	128
Serbian	166
Shor see Non-Slavic Languages	128
Sindhi	168
Sinhalese	170
Sonthal see Santali	164
Syriac (in Cyrillic script) see Non-Slavic Languages--Aisor	115
Syriac (in Hebrew script) see Hebrew and Yiddish	52
Tabasaran see Non-Slavic Languages	129
Tajik see Non-Slavic Languages	129
Tamil	172
Tat see Non-Slavic Languages	129
Tatar see Non-Slavic Languages	129
Tegulu	174
Thai	176
Tibetan	192
Tigrinya	194
Tua Tham see Pali	141
Turkish see Ottoman Turkish	134
Turkmen see Non-Slavic Languages	129
Tuva see Non-Slavic Languages	130
Udekhe see Non-Slavic Languages	130
Udmurt see Non-Slavic Languages	130
Uighur (in Arabic script)	196
Uighur (in Cyrillic script) see Non-Slavic Languages	130
Ukrainian	200
Urdu	202
Uzbek see Non-Slavic Languages	130
White Russian see Belorussian	18
Xanty see Non-Slavic Languages--Khanty	122
Yakut see Non-Slavic Languages	130
Yiddish see Hebrew and Yiddish	52
Yult see Non-Slavic Languages--Eskimo	119